Btech Project - Progress Diary

Arka Sadhu

September 7, 2017

Contents

1 Daily Progress

1.1 Week 1 : July 17 - July 23

1.1.1 July 17 : Monday

Accomplished :

- Met prof and discussed basic project about deep learning and image co-segmentation.
- Prof asked to mail and contact sayan banerjee.

1.1.2 July 18 : Tuesday

Accomplished :

- Mailed Sayan banerjee, but seems like he has a fever, and also didn't give/suggest any papers unfortunately.
- Only saw one image cosegmentation paper by Inria, but didn't have the time to read it.

1.1.3 July 19 : Wednesday

Accomplished :

- Dowloaded a few papers to read.
- Init git repository and added basic work report format (from viterbi internship project).

1.1.4 July 20 : Thursday

Target :

- Try to meet up with Prof/ Sayan and discuss about possible extensions.
- Brainstorm ideas as and when time permits.
- Read as many papers as possible. Complete in-depth reading is not required, simple reading of abstract and reading the relevant papers should be enough.

Accomplished :

- Read the 3 papers :
    - Discriminative clustering for image co-segmentation [1]: abstract gives basic definition of image co-segmentation as well as says unsupervised segmentation of image into foreground and background regions is still a challenging task. Should be worth reading. Images are also as requried.
    - Automatic Image Co-segmentation using geometric saliency [2]: says co-labelling of multiple images is a complex process, instead suggests segmentation on individual images but based on a saliency map obtained by fusing saliency maps of groups of similar images. Images are also as required. Since it is anyways small, should be worth reading.
    - Unsupervised 3D shape segementation and co-segmentation via deep learning [3]: should have been the most related but turns out they do something quite different (at least from first look). They try to automatically segment a single 3D shape or co-segment family of 3D shape. For this they use pre-decompose 3D shape into primitive patches to compute various low level features, then learn high level features in an unsupervised style from low level features based on deep learning, and finally either segmentation or co-segmentation results are fot by patch clustering in high level

feature space. The input images are 3D and not 2D as originally required, as such the paper is on Comptuer graphics apparently. But it should be interesting to read the deep learning part of it, i.e. how did they learn the high level features in the unsupervised style from low level features.

1.1.5 July 21 : Friday

No work done.

1.1.6 July 22 : Saturday

No work done

1.1.7 July 23 : Sunday

No work done.

1.1.8 July 24 : Monday

Meeting with Sayan. Discussion about some basic problems that could be solved. Apparently Video Co-segmentation using Deep learning is not yet looked upon. May be a good breakthrough.

1.1.9 July 25 : Tuesday

Target :

- Read the three papers. [4] , [5], [6]

Accomplished :

- Able to read only 1 paper properly (not too properly).

1.1.10 July 26 : Wednesday

- Couldn't do much. Need to do paper review at the very first.
- Need to do pytorch tutorials as well.
- Paper review to be done in a separate file.

1.1.11 July 27 : Thursday

Accomplished :

- Completed reading all the papers and joting down main points of the paper.
- Meeting with Sayan (main points written below):
  - Video Co-segmentation hasn't been worked on by the Vision community yet.
  - The main aim is to make an end-to-end model using RNN.
  - To circumvent problems of vanishing gradients, we can use LSTM. But for our problem, even LSTM might not suffice.
  - Hence we can use Memory Augmented LSTM, which should be able to take care of the vanishing gradients.

- There is no big datasets for the purpose of Video Co-segmentation. Hence we can look at Weakly unsupervised learning techniques to circumvent this problem.

- Need to read 5 papers at the earliest (by Monday). [7], [8], [9], [10], [11]

1.1.12 July 28 : Friday

No work done

1.1.13 July 29 : Saturday

Probably no work done

2 Week 1 : September

2.1 Sept 6 : Wednesday

Discussion with the prof. Main points noted below. Also has what all needs to be done before tomorrow's meeting.

- Graph Signal. Needs a more concrete definition. Can the vectors be of different dimensions? (I think not, but need to verify this.)

- An image may have 100-200 superpixels, which is not a constant. So can't use the same graph for all of them. Can we exploit the graph structure? GCN though requires the same graph structure to be followed, otherwise it won't be able to exploit the weights and would essentially require new training and generally useless.

- We would also like to encode the spatial connections of the superpixels.

- When we have arbitrary shape, we don't want to have any kind of orientation information, may mess things up.

- Each superpixel would give some form of a feature vector and we would like to use this feature vector as inputs to the GCN. But need to check if the GCN would actually be of any use?

- We want to use Graph structure to get some kind of context between the superpixels. So this would be a context sensitive classifier. Need to do extensive literature survey on this.

- For the case of Co-segmentation the problem is that can we even preserve the graph structure? The nodes? The adjacency graph / matrix. Need to think about these things more carefully.

- Co-segmentation in general may include both seen/unseen categorization as well. We might need to train on one set of classes and still get outputs on another unseen set of classes. Prof gave the same example as mine, 30C2 combinations, or possibly 25C2 and test on the rest 5C2. We can get the corresponding ground truths as well.

- Two aspects to the problem of Co-segmentation. Say G1 has m nodes, and G2 has n nodes. We would like to perform subgraph matching problem. What Avik has done is used a greedy algorithm which gives the results quite fast. But I don't think that would be enough actually. Prof suggested to use Graph Neural Networks to solve this problem. This is actually a combinatorial problem. We would like to learn to solve this combinatorial problem.

- In one case the problem is one-to-one, but here it is possible that it is many-to-many. Because one superpixel may match to two smaller super pixels from the other graph and vice-versa. Solution exists for the one-to-one problem. Sayan's talk tomorrow on the jigsaw puzzle thing is also one-to-one. May need to check the feasibility of such problems.

- Subgraph Matching Using Deep Learning.

- The other aspect is the features that are bein given as inputs. Currently it is hog and color features which are in some sense hand-crafted. But what we probably need is more general features. I am thinking in the direction of dilated convolution. May be that can help. [Need to read this one too.]

- In graphs, the sub-graph matching is NP-complete. To get the MCS : currently greedy algorithm being employed.

- Another thing to note is that currently we are doing only 0th order. We would like it to be of say some k-ordered, i.e. see till k-hops.

- So the problem now effectively becomes matching a group of nodes to a group of nodes. A problem that could arise is that number of nodes of the two matching things may not be exactly the same. What we can possibly do is to bring the number of nodes to be the same, by clustering some very similar nodes together.

2.2 Sept 7 : Thursday

Minutes of the discussion with professor and my answers to the doubts we previously had.

- First if the graph structure is approximately the same, we can simply do an adjacency approximation and say that the learned parameters are shared across the graphs.

- Second if the dimensionality of the vectors in each node is different can go to some fixed length representation using attention mechanism.

- Third is that if the graph is completely different then we can pass on the laplacian as well. My suggestion and partly intuition is that given a graph signal x_{G1} and graph signal x_{G2} belonging to two different graphs, we should be able to find some kind of conversion using the attention mechanism previously described.

- Then we go to Sayan's Jigsaw puzzle. The problem is formulated in RNN way with LSTM cells.

- Need to understand how to apply the graph networks to the subgraph matching problems clearly.

- Also need to understand seq2seq models. One thing I need to know is to how do we manage the problem of having different length vectors in the seq2seq models.

- Need to read about : Graph Neural Networks as to how they are different from graph convolution networks in what way and all. Second is to read the gated graph attention mechanism given for alleviating different graph structure problem.

References

[1] A. Joulin, F. Bach, and J. Ponce, "Discriminative clustering for image co-segmentation," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1943–1950, June 2010.

[2] K. R. Jerripothula, J. Cai, F. Meng, and J. Yuan, "Automatic image co-segmentation using geometric mean saliency," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 3277–3281, Oct 2014.

[3] Z. Shu, C. Qi, S. Xin, C. Hu, L. Wang, Y. Zhang, and L. Liu, "Unsupervised 3d shape segmentation and co-segmentation via deep learning," *Computer Aided Geometric Design*, vol. 43, pp. 39 – 52, 2016. Geometric Modeling and Processing 2016.

[4] H. Fu, D. Xu, B. Zhang, S. Lin, and R. K. Ward, "Object-based multiple foreground video co-segmentation via multi-state selection graph," *IEEE Transactions on Image Processing*, vol. 24, pp. 3415–3424, Nov 2015.

[5] K. Li, J. Zhang, and W. Tao, "Unsupervised co-segmentation for indefinite number of common foreground objects," *IEEE Transactions on Image Processing*, vol. 25, pp. 1898–1909, April 2016.

[6] D. Zhang, O. Javed, and M. Shah, *Video Object Co-segmentation by Regulated Maximum Weight Cliques*, pp. 551–566. Cham: Springer International Publishing, 2014.

[7] M. Siam, S. Valipour, M. Jägersand, and N. Ray, "Convolutional gated recurrent networks for video segmentation," *CoRR*, vol. abs/1611.05435, 2016.

[8] A. Jain, A. R. Zamir, S. Savarese, and A. Saxena, "Structural-rnn: Deep learning on spatio-temporal graphs," *CoRR*, vol. abs/1511.05298, 2015.

[9] M. S. Pavel, H. Schulz, and S. Behnke, "Object class segmentation of rgb-d video using recurrent convolutional neural networks," *Neural Networks*, vol. 88, pp. 105 – 113, 2017.

[10] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," *CoRR*, vol. abs/1505.04366, 2015.

[11] Ç. Gülçehre, S. Chandar, and Y. Bengio, "Memory augmented neural networks with wormhole connections," *CoRR*, vol. abs/1701.08718, 2017.